بسمه تعالی

تمرین سری هفتم درس یادگیری عمیق

پوریا محمدی نسب

۴۰۰۷۲۲۱۳۸

فهرست

سوال ۱ ..... ۳

سوال ۲ ..... ۵

الف ۵

ب ۵

سوال ۳ ..... ۷

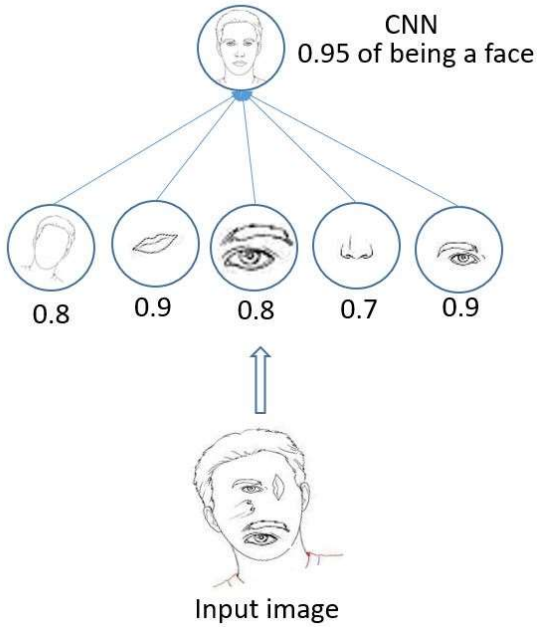
سوال ۴ ..... ۸

References ۱۱

سوال ۱.

از جمله شبکه هایی که در رقابت با شبکه های همگشتی ارایه شدند، شبکه های کپسولی هستند. با مطالعه مقاله CapsNet (Dynamic Routing Between Capsules) توضیح دهید که کپسول چیست و این نوع شبکه ها را با شبکه های همگشتی مقایسه کنید و توضیح دهید که برای رفع کدام محدودیت های شبکه های همگشتی ارایه شده اند.(برای پاسخ دادن به این سوال مطالعه قسمت های ابتدایی مقاله و بررسی اجمالی معماری آن کافی است اما مطالعه این مقاله کاربردی توصیه میشود.)

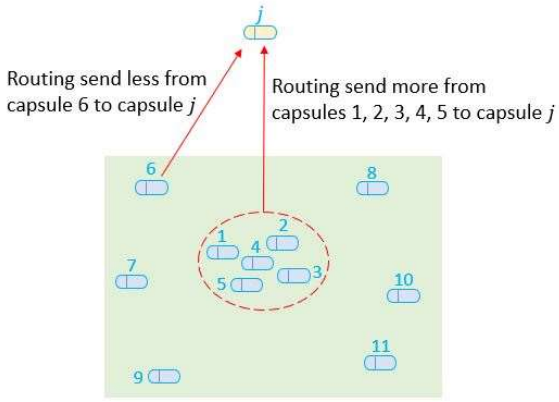
اصل ایده ی CapsNet این است که به structure ساده standard neural networks غنای بیشتری اضافه کنیم. تفاوت اصلی شبکه های همگشتی نیز با شبکه های CapsNet نیز همین مورد است به طوری که در شبکه های همگشتی یک نرون به تنهایی فقط یک عدد را (منفی، مثبت یا صفر) نمایندگی میکند اما مفهوم Capsule این است که گروهی از نورون ها را با هم دسته کنیم و به صورت یک واحد بزرگ تر به آن نگاه کنیم. در شبکه های Capsule به جای ارتباط دو نرون تکی با هم Capsule ها با هم ارتباط دارند. یک Capsule حاوی دو نوع اطلاعات است. ۱) آیا شی مورد نظر را مشاهده کرده است؟ که به این قسمت اصطلاحا presence میگویند. ۲) Viewing condition را در خود نگه دارد که اصطلاحا به آن instantiation میگویند یعنی المان های داخل کپسول نسبت و ارتباط اجزای داخل تصویر را نگه دارند. قسمت viewing condition برای تشخیص جزئی از object است که object part نامیده میشود و اگر object part های یک شی با الگوی درستی کنار هم قرار نگرفته باشند CapsNet قابلیت تشخیص آن را دارد در حالی که در شکبه های همگشتی این قابلیت وجود نداشت. برای درک بهتر این تفاوت شکل زیر را مشاهده کنید.



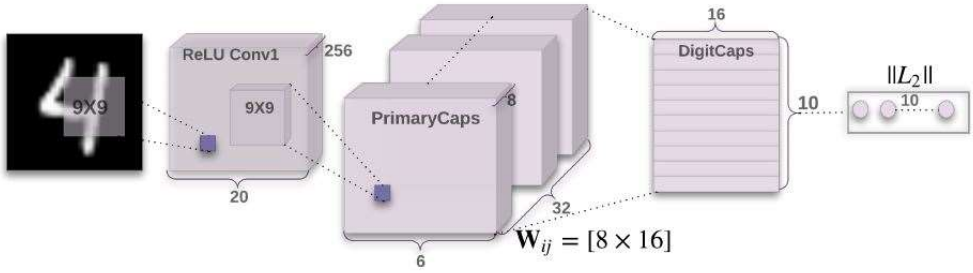
یک شبکه همگشتی با توجه به اینکه توجهی به ارتباط بین object part ها ندارد ورودی تصویر بالا را با احتمال ۰.۹۵ درصد صورت میداند در حالی که اگر به ارتباط اجزا توجه میکرد نباید به این ورودی احتمال بالایی اختصاص میداد. مزیت دیگر

CapsNet ها نسبت به شبکه های همگشتی نیاز کمتر آنها به training data است همچنین روی تصاویر شلوغ نتایج بسیار بهتری دارد دلیل این دو مزیت مفهومی تحت عنوان equivariant است که بدین معناست که در یک تصویر تمام object part ها باید به یک اندازه تغییر کنند.

حال به این موضوع میپردازیم که یک CapsNet چگونه عمل میکند. در هر لایه از شکبه تعداد کپسول وجود دارد. اگر در لایه ی L باشیم و تعدادی کپسول برای المان های یک تصویر داشته باشیم میتوانیم از تکنیکی تحت عنوان Routing استفاده کنیم. این تکنیک مشخص میکند که هر کپسول در لایه ی L چه مقدار در ساختن کپسول سطح بالاتری که در لایه L+1 قرار دارد موثر است.



در شکل بالا مشاهده میکنیم که با استفاده از تکنیک routing ، کپسول های ۱و۲و ۳و ۴و ۵ در ساختن کپسول j در لایه ی بالاتر تاثیر بیشتری دارند اما کپسول ۶ با وزن کمتری به کپسول j متصل است. برای آشنایی بیشتر با ساختار این شبکه ها از تصویر داخل مقاله استفاده کردیم:



برای این مثال در ابتدا یک لایه convolution استفاده میکنیم تا تعدادی feature map اولیه بدست آوریم در ادامه با ترکیب نرون های شبکه primaryCaps را میسازیم که از این قسمت به بعد الگوریتم routing استفاده میشود تا کپسول های لایه های بعدی و سطح بالاتر ساخته شوند. در انتها به کپسول هایی درسطح تشخیص اعداد میرسیم که به راحتی مشخص میکنند که چه عددی در تصویر مشاهده شده است. با توجه به شکل در آخرین لایه ی شامل کپسول ها ۱۰ کپسول داریم که هر کپسول به تنهایی شامل ۱۶ نورون است.

سوال ۲.

به سوالات زیر پاسخ دهید.(منابع خود را ذکر کنید)

الف

مزایا و معایب لایه های ادغام را در شبک ه های همگشتی بیان کنید . آیا استفاده از لایه های ادغام حداکثری و ادغام میانگین در کاربرد و نتیجه حاصل از خروجی با هم تفاوتی دارند؟ اگر پاسخ مثبت است تفاوتشان را بیان کنید.

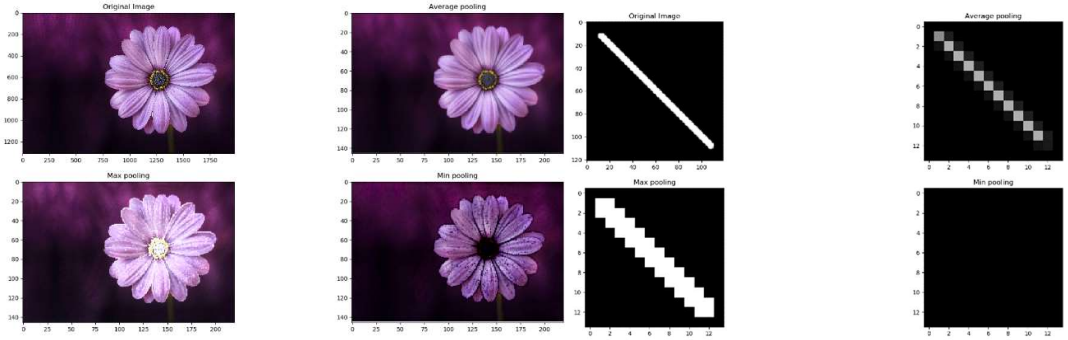
از مزایای لایه ی pooling میتوان به دو مورد زیر اشاره کرد:

۱. باعث کاهش چشمگیر تعداد پارامترهای شبکه میشود که برای قسمت آموزش مدل بسیار مفید است.
۲. بدون نیاز به آموزش پارامتر است پس ریسک اینکه باعث overfit شدن مدل شود را ندارد.
۳. معمولا نتایج نشان میدهد که نتایج بهتری از نظر دقت مدل ارائه میدهد.

در کنار مزایای لایه های pooling تعداد عیب نیز وجود دارد:

۱. لایه های pooling باعث از دست رفتن قسمتی از دیتا میشوند زیرا ذاتا لایه های برگزیننده هستند.
۲. استفاده ی بیش از حد از pooling برخلاف مورد اول مزایا عمل میکند و کاهش بیش از حد تعداد پارامترهای شبکه باعث وقوع underfitting میشود.
۳. نیاز مدل به Hyper parameter های بیشتر مانند سایز pooling و stride.

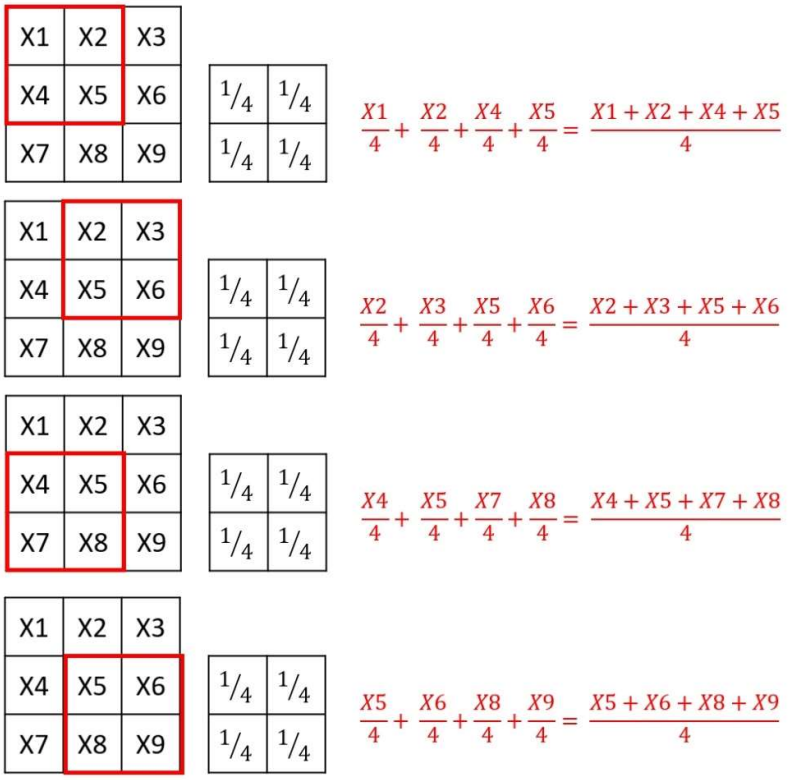
هنگامی که از max pooling با average pooling استفاده کنیم تاثیر چندانی در نتایج مدل حاصل نمیشود. اما در شکل خروجی هنگامی که از average pooling استفاده کنیم خروجی feature map ها ساختاری نرم (smooth) دارند اما وقتی از max pooling استفاده کنیم خروجی معمولا نقاط روشن بیشتری دارد. شکل زیر مقایسه ای از خروجی انواع pooling را نمایش میدهد.



ب

آیا میتوان هر کدام از لایه های ادغام حداکثری و ادغام میانگین را با یک لایه عصبی همگشتی پیاده سازی کرد؟ اگر جواب مثبت است آن را با رسم شکل نشان دهید.

لایه Average pooling با یک لایه کانولوشنی قابل پیاده سازی است. برای این کار کافی است فرض کنید اندازه Kernel مورد نظر N*M باشد، اگر مقادیر هر درایه از کرنل برابر 1/N*M باشد عمل میانگین گیری انجام میشود.



در مورد Max Pooling به دلیل نیاز به تابع Max که یک تابع غیر خطی است نمیتوان این لایه را صرفا با convolution پیاده سازی کرد.

سوال ۳.

شبکه عصبی همگشتی زیر را در نظر بگیرید. در لایه های همگشتی مقادیر به ترتیب برابر با تعداد کانالهای خروجی(تعداد فیلترها)، اندازه فیلتر و تعداد گام ها هستند. فرض کنید ورودی یک تصویر رنگی با اندازه 128 در 128 است. اندازه خروجی و تعداد پارامترها را برای هر لایه به دست آورید.

Conv (64, (5,5), 2)

Conv (64, (3,3), 2)

Max-Pool (3*3)

برای محاسبه شکل خروجی در لایه های Conv داریم:

Output shape = (input size – filter size) / stride + 1 * filter number

همچنین برای محاسبه تعداد پارامترهای این لایه از فرمول زیر استفاده میکنیم:

$$Parameters = ((filter\ size * C_{in}) + 1) * C_{out}$$

برای لایه اول:

Output shape = (128 - 5) / 2 + 1 * 64 = **(62 * 62 * 64)**

Parameters = ((3 * 5 * 5) + 1) * 64 = **4864**

برای لایه دوم:

Output shape = (62 - 3) / 2 + 1 * 64 = **(30 * 30 * 64)**

Parameters = ((64 * 3 * 3) + 1) * 64 = **36928**

برای لایه سوم:

Output shape = (input size / stride) * filter number = 30 / 3 = **(10 * 10 * 64)**

Parameters = **0**

سوال ۴.

ورودی یک لایه همگشتی (X) با ابعاد سه در سه را در نظر بگیرید. فیلتر F با ابعاد دو در دو روی ورودی X اعمال شده است. روی خروجی این لایه همگشتی، یک لایه ادغام میانگین سراسری (GAP) اعمال میشود که خروجی نهایی یک عدد خواهد شد. با توجه به این که گرادیان تابع اتلاف نسبت به این خروجی نهایی که یک عدد است، ۱ میشود، با استفاده از الگوریتم پس انتشار خطا گرادیان های این لایه همگشتی را به دست آورید.

3	4	5
2	1	-3
4	-2	0

X

2	0
-3	1

F

کانوالو کردن X و F:

$$O_{11} = (3*2) + (4*0) + (-3 * 2) + (1*1) = 1$$

$$O_{12} = (2 * 4) + (5 * 0) + (-3 * 1) + (-3 * 1) = 2$$

$$O_{21} = (2 * 2) + (1 * 0) + (-3 * 4) + (-2 * 1) = -10$$

$$O_{22} = (1 * 2) + (-3 * 0) + (-2 * -3) + (0 * 1) = 8$$

1	2
-10	8

O

اعمال GAP روی خروجی O :

$$(1 + 2 - 10 + 8) / 4 = 0.25$$

محاسبه گرادیان:

$$Local\ Gradiant = \frac{\partial O}{\partial F}$$

$$O_{11} = X_{11}F_{11} + X_{12}F_{12} + X_{21}F_{21} + X_{22}F_{22}$$

$$\frac{\partial O_{11}}{\partial F_{11}} = X_{11}, \quad \frac{\partial O_{11}}{\partial F_{12}} = X_{12}, \quad \frac{\partial O_{11}}{\partial F_{21}} = X_{21}, \quad \frac{\partial O_{11}}{\partial F_{22}} = X_{22}$$

به طور مشابه میتوانیم local gradient ها را برای O_{12}، O_{21} و O_{22} نیز حساب کنیم.

حال برای محاسبه $\partial L/\partial F$ از قاعده زنجیری استفاده میکنیم:

$$\frac{\partial L}{\partial F} = \sum_{k=1}^{M} \frac{\partial L}{\partial O_k} \times \frac{\partial O_k}{\partial F}$$

اگر در رابطه ی بالا سیگما را گسترش دهیم:

$$\frac{\partial L}{\partial F_{11}} = \frac{\partial L}{\partial O_{11}} \times \frac{\partial O_{11}}{\partial F_{11}} + \frac{\partial L}{\partial O_{12}} \times \frac{\partial O_{12}}{\partial F_{11}} + \frac{\partial L}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial F_{11}} + \frac{\partial L}{\partial O_{22}} \times \frac{\partial O_{22}}{\partial F_{11}}$$

$$\frac{\partial L}{\partial F_{12}} = \frac{\partial L}{\partial O_{11}} \times \frac{\partial O_{11}}{\partial F_{12}} + \frac{\partial L}{\partial O_{12}} \times \frac{\partial O_{12}}{\partial F_{12}} + \frac{\partial L}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial F_{12}} + \frac{\partial L}{\partial O_{22}} \times \frac{\partial O_{22}}{\partial F_{12}}$$

$$\frac{\partial L}{\partial F_{21}} = \frac{\partial L}{\partial O_{11}} \times \frac{\partial O_{11}}{\partial F_{21}} + \frac{\partial L}{\partial O_{12}} \times \frac{\partial O_{12}}{\partial F_{21}} + \frac{\partial L}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial F_{21}} + \frac{\partial L}{\partial O_{22}} \times \frac{\partial O_{22}}{\partial F_{21}}$$

$$\frac{\partial L}{\partial F_{11}} = \frac{\partial L}{\partial O_{11}} \times \frac{\partial O_{11}}{\partial F_{11}} + \frac{\partial L}{\partial O_{12}} \times \frac{\partial O_{12}}{\partial F_{11}} + \frac{\partial L}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial F_{11}} + \frac{\partial L}{\partial O_{22}} \times \frac{\partial O_{22}}{\partial F_{11}}$$

با جایگذاری گرادیان های محلی داریم:

$$\frac{\partial L}{\partial F_{11}} = \frac{\partial L}{\partial O_{11}} \times X_{11} + \frac{\partial L}{\partial O_{12}} \times X_{12} + \frac{\partial L}{\partial O_{21}} \times X_{21} + \frac{\partial L}{\partial O_{22}} \times X_{22}$$

$$\frac{\partial L}{\partial F_{12}} = \frac{\partial L}{\partial O_{11}} \times X_{12} + \frac{\partial L}{\partial O_{12}} \times X_{13} + \frac{\partial L}{\partial O_{21}} \times X_{22} + \frac{\partial L}{\partial O_{22}} \times X_{23}$$

$$\frac{\partial L}{\partial F_{21}} = \frac{\partial L}{\partial O_{11}} \times X_{21} + \frac{\partial L}{\partial O_{12}} \times X_{22} + \frac{\partial L}{\partial O_{21}} \times X_{31} + \frac{\partial L}{\partial O_{22}} \times X_{32}$$

$$\frac{\partial L}{\partial F_{22}} = \frac{\partial L}{\partial O_{11}} \times X_{22} + \frac{\partial L}{\partial O_{12}} \times X_{23} + \frac{\partial L}{\partial O_{21}} \times X_{32} + \frac{\partial L}{\partial O_{22}} \times X_{33}$$

اگر به روابط بالا توجه کنیم متوجه میشود که این عبارات در واقع حال کانولوشن X با گرادیان های محلی (loss بدست آمده از لایه ی قبلی) است.

در ادامه برای محاسبه $\partial L/\partial X$ داریم:

$$O_{11} = X_{11}F_{11} + X_{12}F_{12} + X_{21}F_{21} + X_{22}F_{22}$$

$$\frac{\partial O_{11}}{\partial X_{11}} = F_{11}, \quad \frac{\partial O_{11}}{\partial X_{12}} = F_{12}, \quad \frac{\partial O_{11}}{\partial X_{21}} = F_{21}, \quad \frac{\partial O_{11}}{\partial X_{22}} = F_{22}$$

برای سایر O ها نیز به همین روش میتوان عمل کرد.

$$\frac{\partial L}{\partial X_i} = \sum_{k=1}^{M} \frac{\partial L}{\partial O_k} \times \frac{\partial O_k}{\partial X_i}$$

$$\frac{\partial L}{\partial X_{11}} = \frac{\partial L}{\partial O_{11}} \times F_{11}$$

$$\frac{\partial L}{\partial X_{12}} = \frac{\partial L}{\partial O_{11}} \times F_{12} + \frac{\partial L}{\partial O_{12}} \times F_{11}$$

$$\frac{\partial L}{\partial X_{13}} = \frac{\partial L}{\partial O_{12}} \times F_{12}$$

$$\frac{\partial L}{\partial X_{21}} = \frac{\partial L}{\partial O_{11}} \times F_{21} + \frac{\partial L}{\partial O_{21}} \times F_{11}$$

$$\frac{\partial L}{\partial X_{22}} = \frac{\partial L}{\partial O_{11}} \times F_{22} + \frac{\partial L}{\partial O_{12}} \times F_{21} + \frac{\partial L}{\partial O_{21}} \times F_{12} + \frac{\partial L}{\partial O_{22}} \times F_{11}$$

$$\frac{\partial L}{\partial X_{23}} = \frac{\partial L}{\partial O_{12}} \times F_{22} + \frac{\partial L}{\partial O_{22}} \times F_{12}$$

$$\frac{\partial L}{\partial X_{31}} = \frac{\partial L}{\partial O_{21}} \times F_{21}$$

$$\frac{\partial L}{\partial X_{32}} = \frac{\partial L}{\partial O_{21}} \times F_{22} + \frac{\partial L}{\partial O_{22}} \times F_{21}$$

$$\frac{\partial L}{\partial X_{33}} = \frac{\partial L}{\partial O_{22}} \times F_{22}$$

References

1) https://www.techopedia.com/definition/33216/capsule-network-capsnet

2) https://www.researchgate.net/publication/341870683_CapsNets_algorithm

3) https://www.geeksforgeeks.org/cnn-introduction-to-pooling-layer/

4) www.quora.com/Are-there-any-weaknesses-in-the-use-of-max-pooling-and-average-pooling

5) https://deepdatascience.wordpress.com/2017/02/09/pooling-intro-adv-and-disadvantage/

6) medium.comwhich-pooling-method-is-better-maxpooling-vs-minpooling-vs-average-pooling

7) https://pavisj.medium.com/convolutions-and-backpropagations-46026a8f5d2c

8) https://towardsdatascience.com/backpropagation-in-a-convolutional-layer-24c8d64d8509